Supplementary Information

Title: A prehistoric East-Asian *Yersinia pestis* genome and a ~5.3 ka trans-Eurasian expansion of plague

Bing Sun[1]#, Yarong Wu[2]#, Daxuan Zhang[1]#, Kai Mu[2]#, Yan Guo[2], Shuang Fu[1], Mingru Sui[1], Chunxiang Li[1], Wenrui Zhang[3], Shan Chen[4], Andrei Poliakov[5], Igor Lazaretov[5], Qiankun Quan[6], Yajun Song[2], Ruifu Yang[2], Yinqiu Cui[1,7]*, Yujun Cui[2]*

1. Key Laboratory for Paleogenetics and Molecular Evolution of Jilin Province, School of Life Sciences, Jilin University, Changchun, 130012, China
2. State Key Laboratory of Pathogen and Biosecurity, Academy of Military Medical Sciences, Beijing, 100071, China
3. Hebei Provincial Institute of Cultural Heritage and Archaeology, Shijiazhuang, 050011, China
4. School of History, Liaoning University, Shenyang, 110136, China
5. Institute for the History of Material Culture of the Russian Academy of Sciences, St. Petersburg, 191181, Russia
6. School of Archaeology, Jilin University, Changchun, 130012, China
7. Key Laboratory for Molecular Enzymology and Engineering of Ministry of Education, School of Life Sciences, Jilin University, Changchun, 130012, China

# These authors contributed equally: Bing Sun, Yarong Wu, Daxuan Zhang, Kai Mu.

* Corresponding author. Email: cuiyujun.new@gmail.com (Yujun Cui); cuiyq@jlu.edu.cn (Yinqiu Cui).

Table of Contents:

Supplementary Notes

Supplementary Note 1: Archaeological background of Zhengjiagou (ZJG) site

The ZJG site (40°58′15.6″N, 115°16′48″E) is located approximately 100 meters west of Zhengjiagou Village in Xuanhua District, Zhangjiakou, Hebei Province, China. Situated in hilly terrain within the Xiong'er Mountain extension of the Yin Mountain Range, the site lies at an elevation of 1,014.6 meters above sea level. Radiocarbon dating of Stone Cairn No. 1 indicates an occupation period spanning approximately 5,100 to 4,800 BP.

Excavations revealed a three-tiered, stone-platform burial complex accompanied by over 60 tombs and 150 ritual features, yielding more than 200 artifacts, including jade ornaments, painted pottery, and shell-adorned objects. Architecturally, the site features three ascending, stone-reinforced earthen terraces enclosed by perimeter walls, spatially divided into centralized burial zones and ritual precincts, reflecting a deliberate spatial hierarchy and social stratification. Burial practices predominantly involved secondary burials placed within stone coffins or cists, accompanied by ritual assemblages of amazonite beads, ceremonial jade rings, and other prestige goods.

While the burial type, ritual pits, and jade artifacts share features characteristic of the Hongshan Culture (circa 4,700-2,900 BP), the ZJG assemblage exhibits unique regional traits. The coexistence of Hongshan-inspired monumental architecture and distinctive decorative repertoires underscores the ZJG site's role as a cultural interface, potentially mediating exchanges between the Hongshan core and peripheral groups. Individual CY470 was recovered from the ZJG site (Fig. 1c). The burial contained nine funerary goods, including two jade axes, one jade artifact, one jade bead, and five pieces of stones.

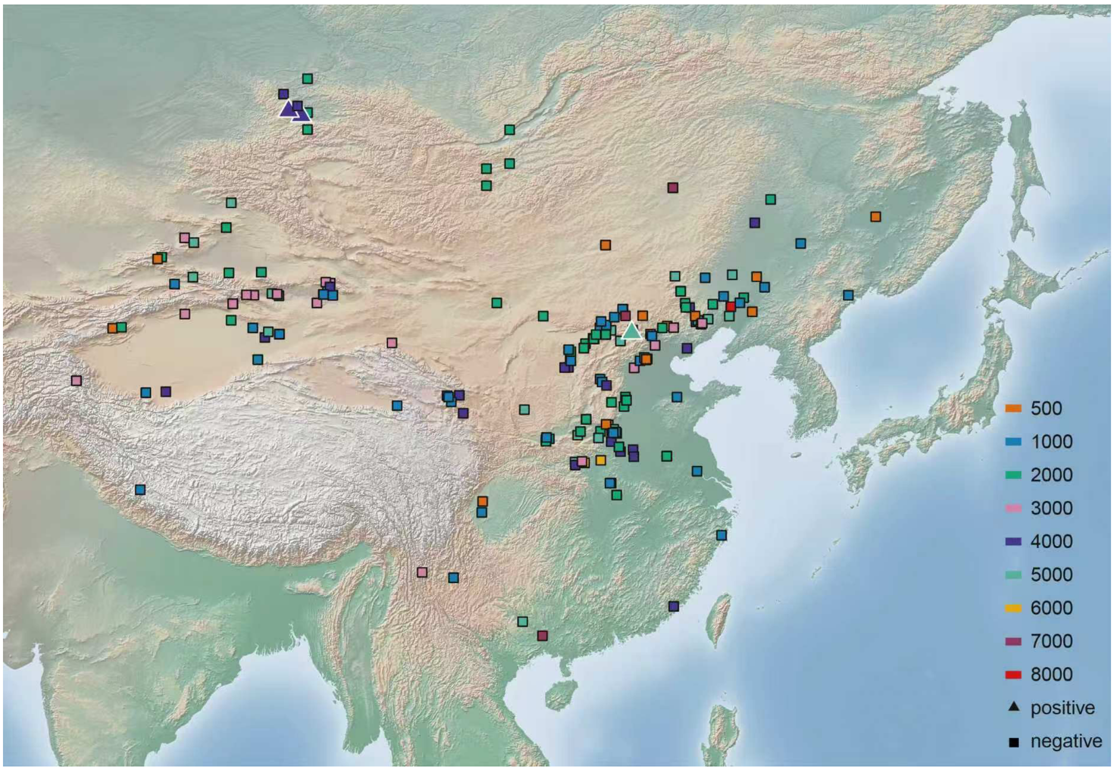
Supplementary Note 2: Archaeological background of the Minusinsk Basin, Russia

The Okunev culture represents a significant Early Bronze Age archaeological complex in the Minusinsk Basin of the Middle Yenisei River region, Russia[1]. Radiocarbon AMS dating places its period between the 26th and 18th centuries BCE, spanning approximately nine centuries. This culture is distinguished by its unique mortuary architecture, notably collective burial grounds featuring stone enclosures[2].

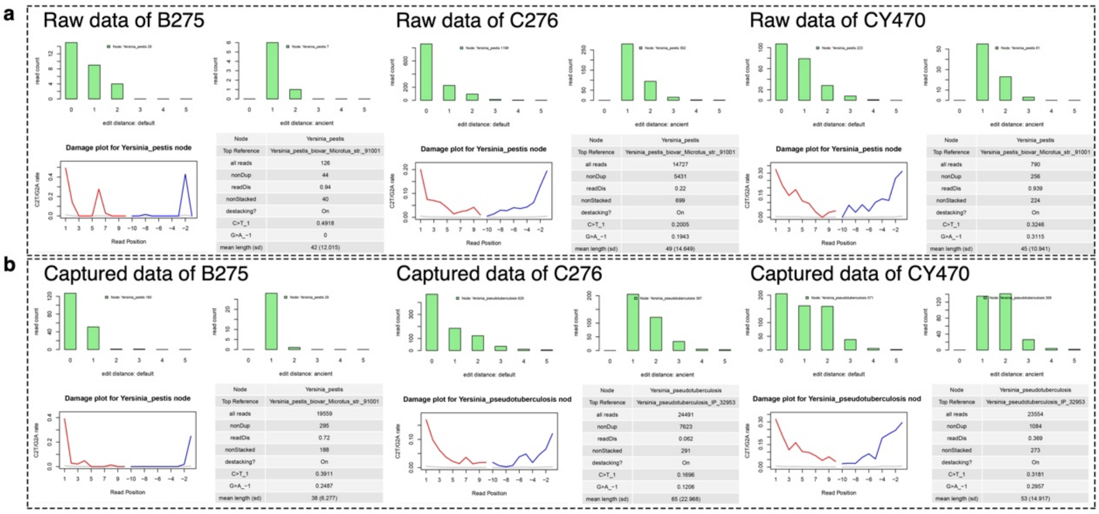
The individual B275 was recovered from the Grave 11 at the Uybat-Charkov cemetery[3], located at the northwest of Charkov Village in the Republic of Khakassia (Fig. 1c). Five radiocarbon dates from two laboratories, obtained from multiple burial materials, consistently date the grave to the 25th-24th centuries BC. Combined with the burial mound's design and radiocarbon dates, this assigns it to the Uybat stage of the Okunev culture. The grave 11, situated in the southwestern sector of the kurgan, features a rectangular pit oriented along a northwest-southeast axis. The upper chamber measures 2 × 1.35 meters in area and 0.5 meters in depth. A secondary excavation at its base extends to 1.4 × 0.95 meters and 0.7 meters in depth, resulting in a total pit depth of 1.2 meters. The burial pit was covered with 5-6 layers of sandstone slabs at the continental shoulder level, supported by horizontal wooden planks. Within the pit, there are two individuals. B275 corresponds to Skeleton 2, a 12-14-year-old boy, who was positioned in a supine posture near the southwestern wall with his head oriented to the northwest. A perforated white stone sphere was found adjacent to the right ilium of his pelvis, and a jar-shaped vessel was placed at his feet.

Individual C276 was recovered from Grave 6 of Kurgan 14 at the Itkol-II site, located in the Shirinsky District of the Republic of Khakassia, 3.5 km east of the village center of Shira (Fig. 1c). The site dates to the early Uybat stage of the Okunev culture (24th-23rd centuries BC). The grave contained an undisturbed burial of a female aged 13-15 years, accompanied solely by a small copper awl with a wooden handle. The deep burial pit featured a stone-slab ceiling (three layers) resting on ledges, with traces of wedging tiles from a missing stele above—a unique ritual element recorded thus far.

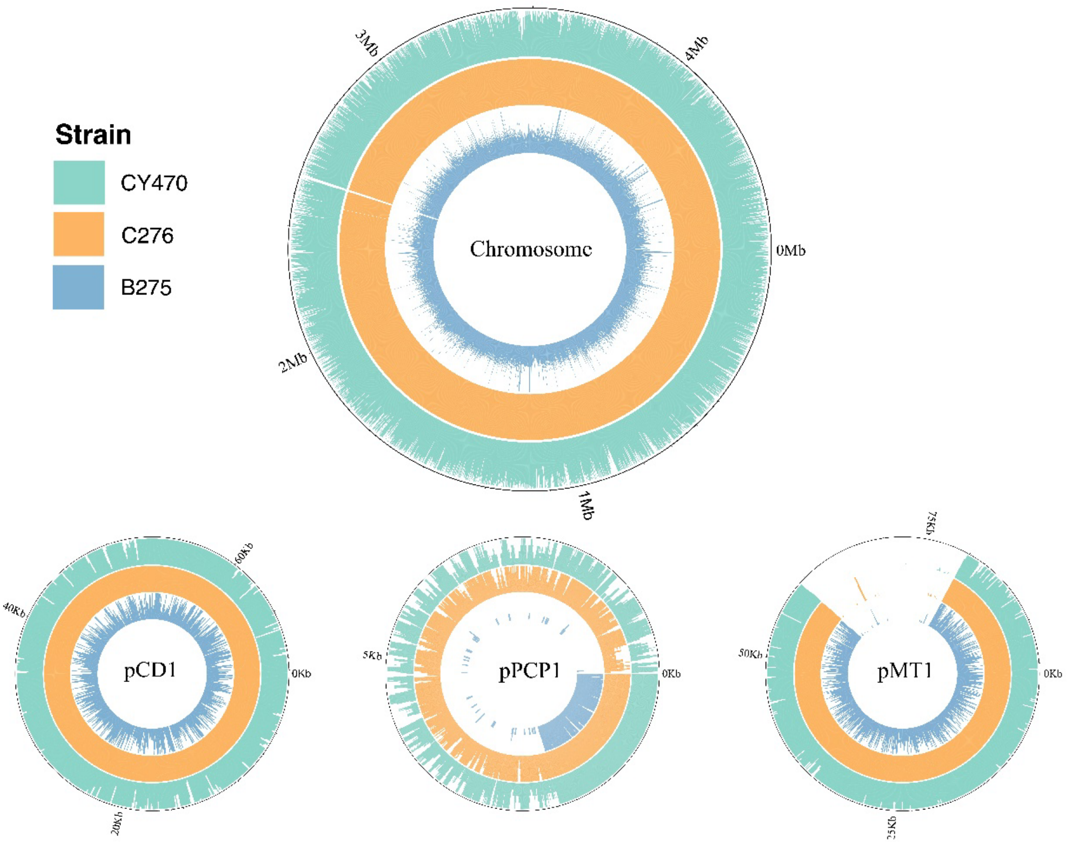
Supplementary Figures



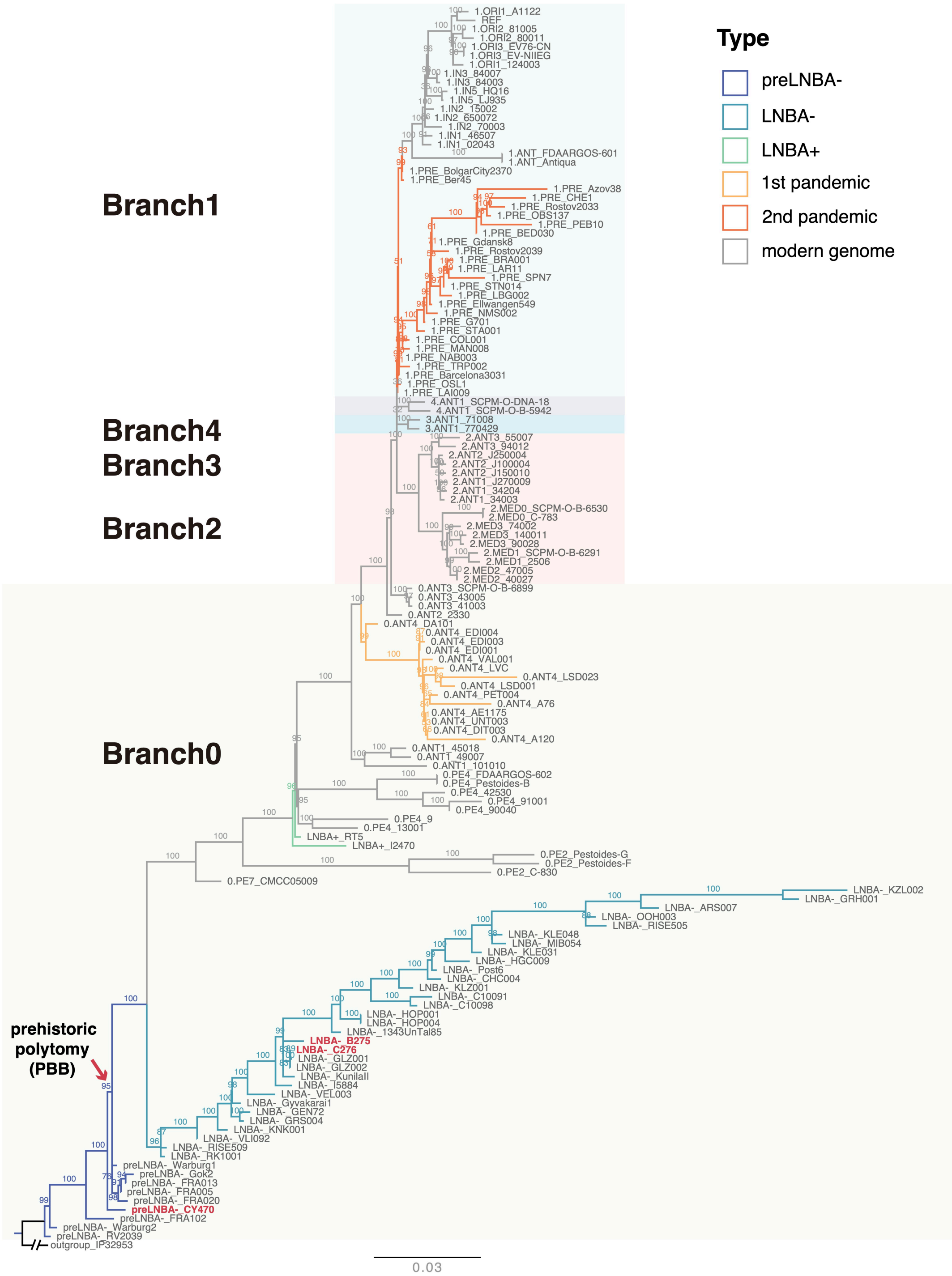
Supplementary Fig. 1 Spatiotemporal distribution of ancient *Y. pestis*-screened genomes. Geographic distribution of *Y. pestis*-positive and -negative ancient genomes analyzed in this study across different time periods. Each point represents an archaeological site, with point color indicating its temporal attribution. Populations in Xinjiang were broadly influenced by the western Eurasian steppe, so Xinjiang groups are classified primarily by archaeological chronology into three periods: Bronze Age (5,000-3,000 BP), Iron Age (3,000-2,000 BP), and Historical Era (2,000-0 BP). In the Central Plain of East Asia over the past ~8,000 years, the main cultural horizons are as follows: Peiligang culture (8,500-7,000 BP), Yangshao culture (7,000-5,000 BP), Miaodigou culture (a transitional phase from Yangshao to Longshan, 5,000-4,300 BP), Longshan culture (4,300-3,800 BP), Erlitou culture (3,800-3,600 BP), Shang dynasty (3,600-3,046 BP), Zhou dynasty (1,046-2,256 BP), Qin dynasty (2,221-2,207 BP), Han dynasty (2,202-1,780 BP), and subsequent dynasties in the historical period.



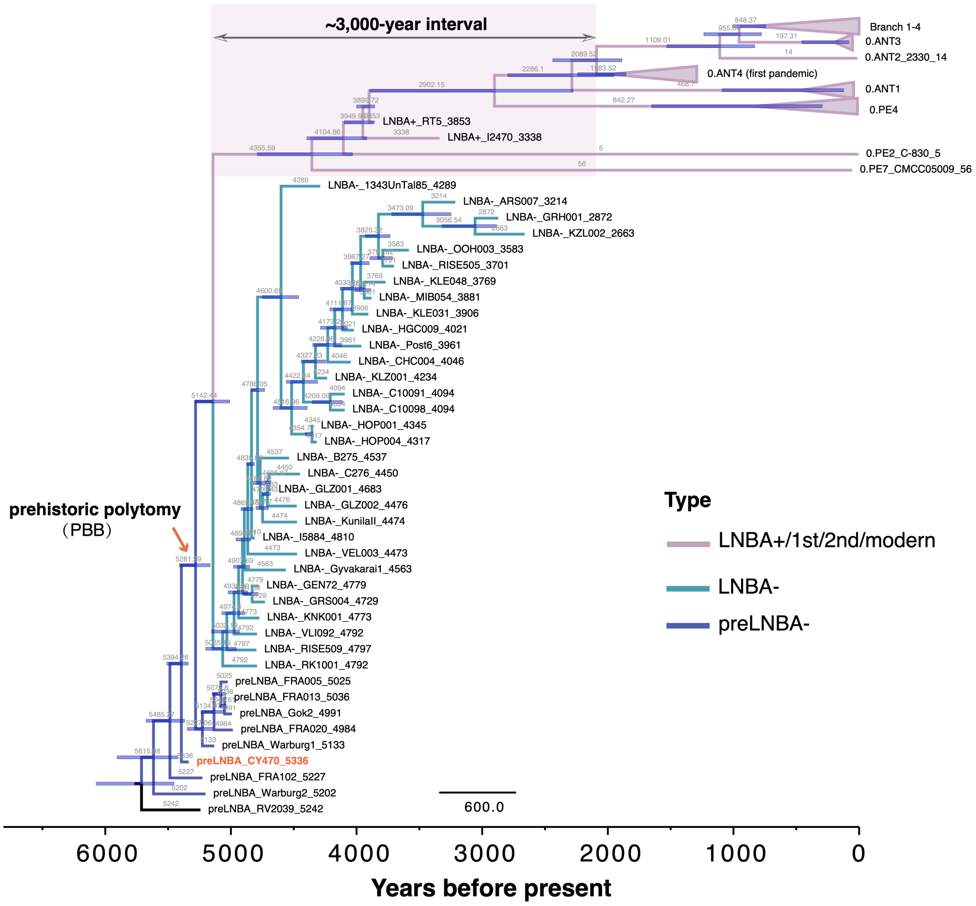
Supplementary Fig. 2 HOPS analysis of ancient *Y. pestis*-positive samples before and after enrichment. a, Pre-enrichment HOPS results for samples B275, C276, and CY470. Upper panels show edit distance distributions with clear peaks at low values. Lower panels display characteristic ancient DNA damage patterns in *Y. pestis*-assigned reads. **b,** Post-enrichment results for samples B275, C276, and CY470. Edit distance profiles and damage patterns remain consistent with pre-enrichment data.



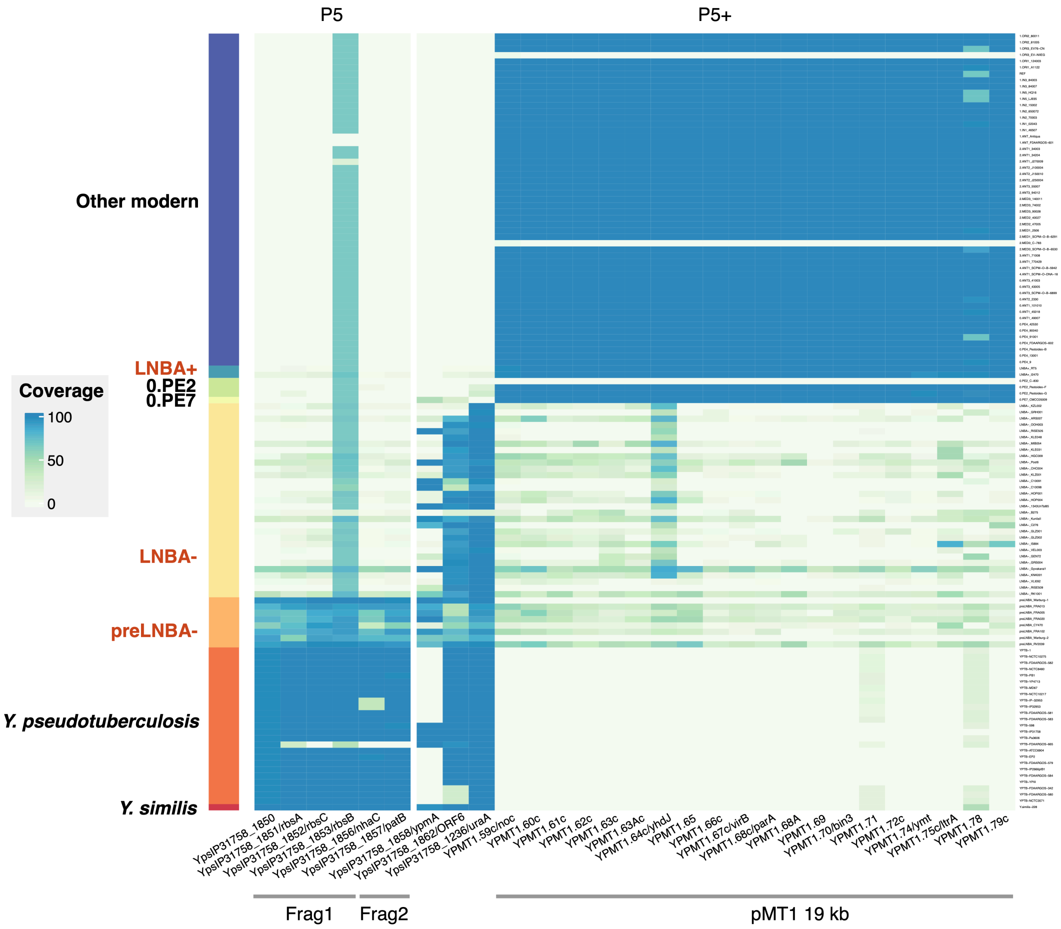
Supplementary Fig. 3 Chromosomal and plasmid coverage plots generated using Circos. Coverage depth was capped at 15× and calculated in 1,000-bp windows for the chromosome, 100-bp windows for the pCD1 and pMT1 plasmids, and 1-bp windows for the pPCP1 plasmid. From outermost to innermost rings: CY470 (green), C276 (yellow), and B275 (blue). Sample B275 exhibits partial pPCP1 coverage (~25% at >1×) and lacks the core virulence locus *pla*, likely owing to lower pPCP1 copy number or preferential DNA degradation in this burial context, as well as exclusion of pPCP1 from enrichment to avoid interference with the capture of the other two plasmids. Source data are provided as a Source Data file.



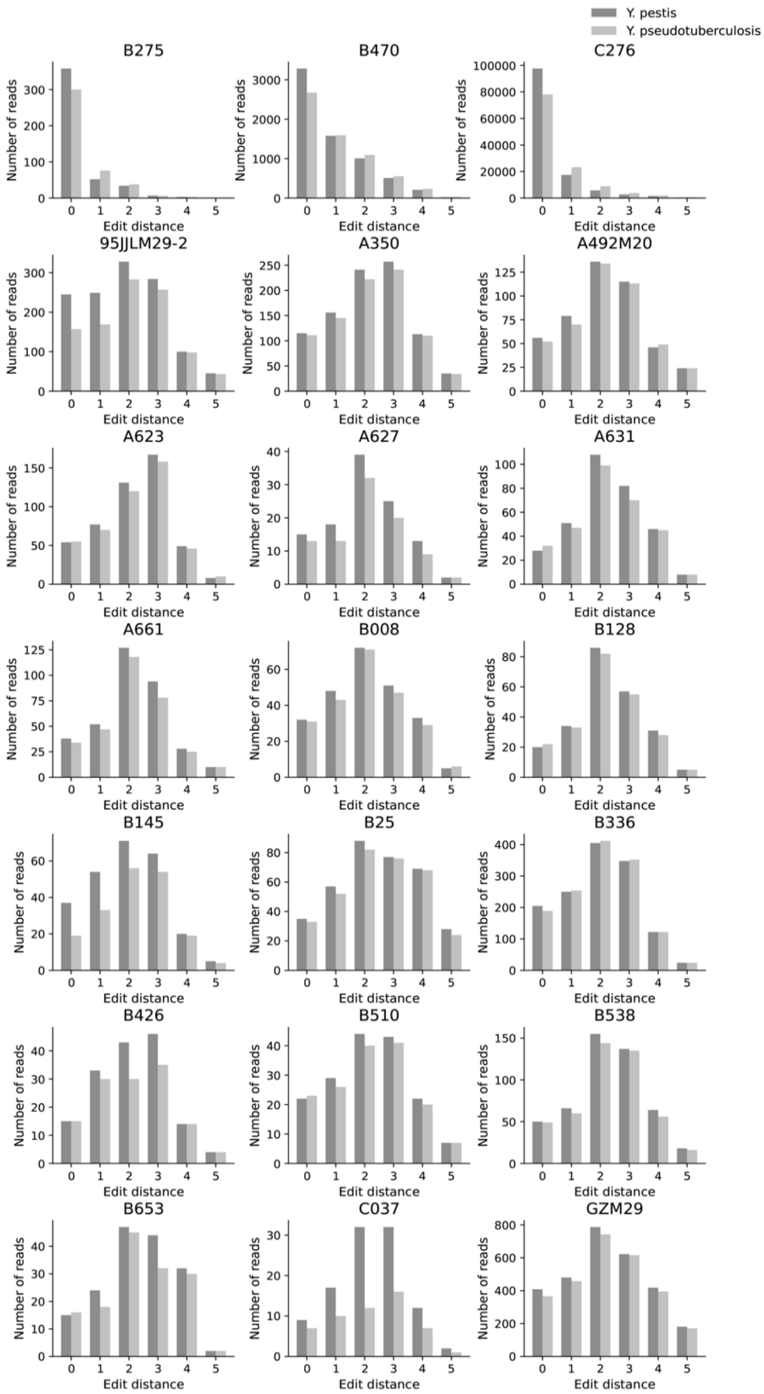
Supplementary Fig. 4 Maximum-likelihood phylogeny of *Y. pestis*. The topology is consistent with that shown in Fig. 1d, but here presented as the fully resolved, uncollapsed tree. The five major branches (branches 0-4) are highlighted with distinct colored shading, and the prehistoric polytomy is marked with a red arrow. Terminal nodes are labeled with phylogroup and strain, formatted as "phylogroup_strain".



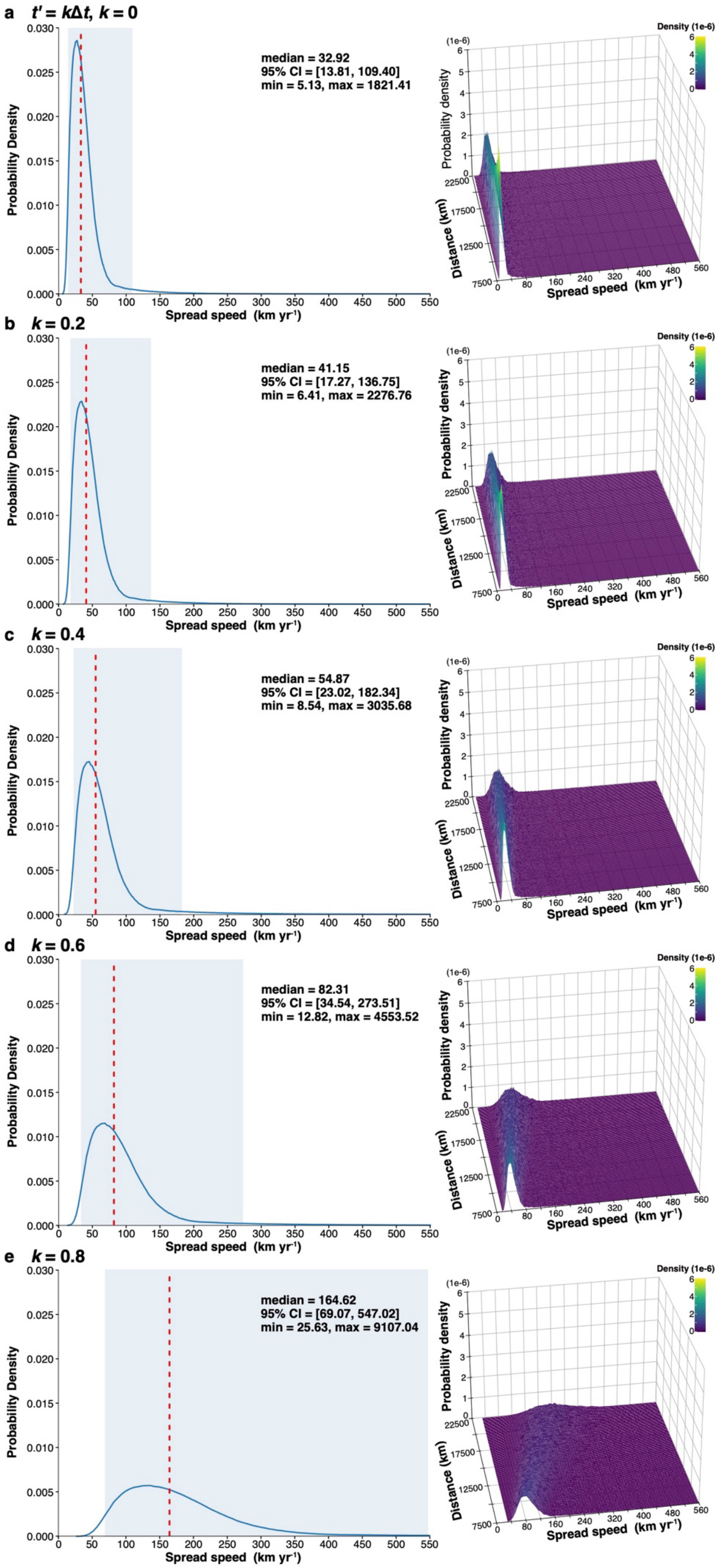
Supplementary Fig. 5 Molecular dating of *Y. pestis* using BEAST. Estimation of the time to the most recent common ancestor (tMRCA) across *Y. pestis*. Except for three modern genomes lacking documented sampling times, the dataset comprises the remaining 136 strains from the total of 139 genomes. Branch lengths are scaled to sample ages, and the median estimated root age of *Y. pestis* is 5.71 kyr (95% highest posterior density (HPD): 5.45-6.07 kyr). The timescale is shown in years before present, where "present" corresponds to the most recently sampled modern strain (2017). Some clades are collapsed for clarity. The ~3,000-year interval spans the diversification of the LNBA$^+$ lineage (~5,142 BP; the "twins" stem splitting into the *ymt*-negative LNBA$^-$ and *ymt*-positive LNBA$^+$ clades) and the tMRCA of the First Pandemic lineage (~1,984 BP). Genomic evidence from human remains during this period is rare; only two LNBA$^+$ genomes, RT5 (3,868-3,704 cal. BP) and I2470 (3,361-3,181 cal. BP), have been reported to date.



Supplementary Fig. 6 Identification of large-fragment gain and loss events in prehistoric *Y. pestis*. The heatmap shows strains as rows and gene gains or losses as columns. Genes contiguous within the same deletion fragment are connected by black horizontal lines below the heatmap. The left color bar indicates the phylogroups of *Y. pestis*, with ancient genome phylogroups labelled in red text. Presence and absence of these fragments are also shown for *Y. pseudotuberculosis* and *Y. similis*. Heatmap colors represent coverage of mapped reads for the corresponding genes. The loss of a 19-kb pMT1 region in three modern *Y. pestis* genomes is due to the lack of short-read sequencing data and assemblies containing only the chromosome. Source data are provided as a Source Data file.

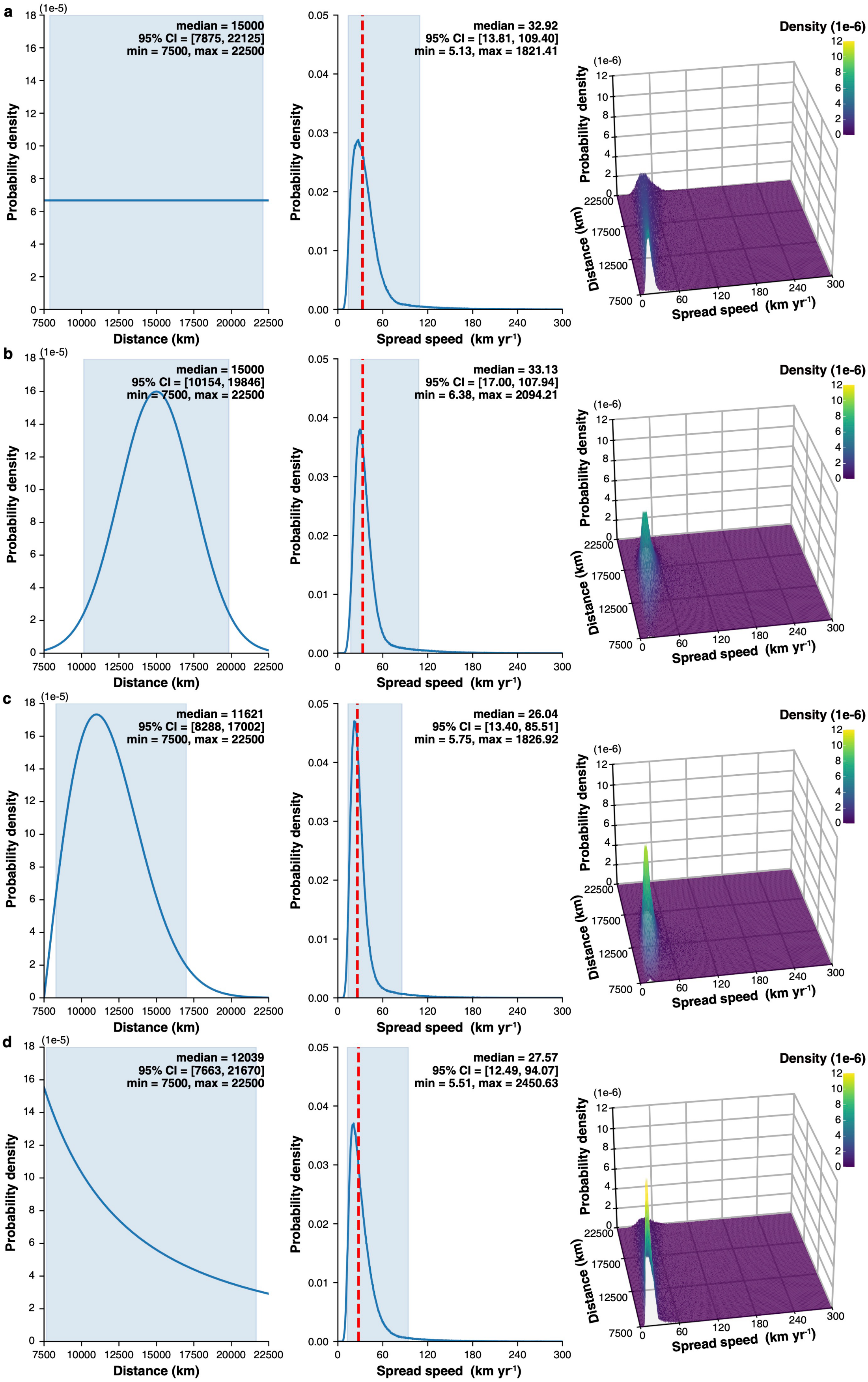


Supplementary Fig. 7 Distribution of edit distances for 21 samples (competitive alignment score > 0) screened from 1400 total samples. Dark gray bars show the edit distance of reads mapped to the *Y. pestis* reference genome, and light gray bars show the edit distance of reads mapped to the *Y. pseudotuberculosis* reference genome. The three positive samples (B275, CY470, and C276), located in the top row of panels, exhibit a clear decreasing trend in read counts with increasing edit distance, whereas the remaining samples do not show this trend. Source data are provided as a Source Data file.



a t′ = kΔt, k = 0
median = 32.92
95% CI = [13.81, 109.40]
min = 5.13, max = 1821.41
b k = 0.2
median = 41.15
95% CI = [17.27, 136.75]
min = 6.41, max = 2276.76
c k = 0.4
median = 54.87
95% CI = [23.02, 182.34]
min = 8.54, max = 3035.68
d k = 0.6
median = 82.31
95% CI = [34.54, 273.51]
min = 12.82, max = 4553.52
e k = 0.8
median = 164.62
95% CI = [69.07, 547.02]
min = 25.63, max = 9107.04
Probability Density
Spread speed (km yr⁻¹)
Probability density
Distance (km)
Density (1e-6)

Supplementary Fig. 8 Delay-adjusted spread speeds of prehistoric *Y. pestis*. **a-e** show increasing non-propagation time $t' = k\Delta t,\ k \in \{0, 0.2, 0.4, 0.6, 0.8\}$. **a,** corresponds to $t' = 0$ and reproduces the baseline results shown in Fig. 2c, d. Left column is the one-dimensional probability density of the delay-adjusted spread speed; the shaded band marks the 95% CI and the red dashed line the median. Right column is the joint density for distance (y-axis, km) and spread speed (x-axis, km yr^{-1}); color gradient encodes density. For comparability, the axis scales are held constant across rows; only the neighborhood around the 95% CI of speed is shown, and the extreme high-speed tail is omitted. All densities are based on two million Monte-Carlo draws. Source data are provided as a Source Data file.


a
median = 15000
95% CI = [7875, 22125]
min = 7500, max = 22500
Probability density
Distance (km)
median = 32.92
95% CI = [13.81, 109.40]
min = 5.13, max = 1821.41
Spread speed (km yr⁻¹)
Density (1e-6)
b
median = 15000
95% CI = [10154, 19846]
min = 7500, max = 22500
median = 33.13
95% CI = [17.00, 107.94]
min = 6.38, max = 2094.21
c
median = 11621
95% CI = [8288, 17002]
min = 7500, max = 22500
median = 26.04
95% CI = [13.40, 85.51]
min = 5.75, max = 1826.92
d
median = 12039
95% CI = [7663, 21670]
min = 7500, max = 22500
median = 27.57
95% CI = [12.49, 94.07]
min = 5.51, max = 2450.63

Supplementary Fig. 9 Sensitivity of inferred spread speeds to distance priors. a-d compare four alternative priors for geographic distance s on the common support $[7500, 22500]$ km: **a**, $s \sim \mathrm{Uniform}(7500{,}22500)$; **b**, $s \sim \mathcal{N}(15000{,}2500)$ truncated to $[7500, 22500]$; **c**, shifted Rayleigh $s = 7500 + X$ with $X \sim \mathrm{Rayleigh}(\sigma = 3500)$ truncated at 22500; **d**, truncated Inverse Gaussian $s \sim \mathrm{InvGauss}(\mu = 12000, \lambda = 3000)$ on $[7500, 22500]$. Left column shows the one-dimensional prior density of s; the shaded band denotes the 95% CI. Middle column shows the one-dimensional density of the inferred spread speed V ($\mathrm{km\ yr^{-1}}$) obtained by Monte Carlo propagation of uncertainty; the shaded band denotes the 95% CI and the red dashed line the median. Right column shows the joint density of distance (y-axis, km) and spread speed (x-axis, $\mathrm{km\ yr^{-1}}$); the color gradient encodes density. Only the neighborhood around the 95% CI of spread speed is shown (0-300 $\mathrm{km\ yr^{-1}}$), and the extreme high-speed tail is omitted. All densities are based on two million Monte Carlo draws. Source data are provided as a Source Data file.

Supplemental References

1 Poliakov, A. V. & Lazaretov, I. P. Current state of the chronology for the palaeometal period of the Minusinsk basins in southern Siberia. *Journal of Archaeological Science: Reports* **29**, 102125 (2020). https://doi.org/10.1016/j.jasrep.2019.102125

2 Poliakov, A. *Chronology and cultural genesis of the Paleometal epoch sites in Minusinsk basins*. (Sankt-Peterburg: IIMK RAN, 2022).

3 I, L. & A, P. Investigations of the Uibat Charkov Burial Ground and New Data on the Early Stage of the Ocunevskaya Culture Development. *Teoriya i praktika arkheologicheskikh issledovaniy* **3**, 41-69 (2018). https://doi.org/10.14258/tpai(2018)3(23).-03